Assessment of Small Molecule Library against Specific Target(s) in Various Cancers by in silico and in vitro approaches for designing the possible Repositioning Strategies

PhD Synopsis

Scholar	**Mohmad Rafeeq Lone**
Enrolment Number	19163CUKmr002
Date of Registration	05 October 2021
Supervisor	**Dr. Abid Hamid Dar** Associate Professor and Head Department of Biotechnology Central University of Kashmir
Co-supervisors	**Prof. M. Afzal Zargar** Department of Biotechnology Central University of Kashmir
	Dr. Nissar Ahmad Wani Scientist-D Department of Biotechnology Central University of Kashmir
Department	**Biotechnology**
School	**Life Sciences**

Central University of Kashmir

Ganderbal, J&K

Title: Assessment of Small Molecule Library Against Specific Target(s) in Various Cancers by *in silico* and *in vitro* Approaches for Designing the Possible Repositioning Strategies

1. Background

Cancer is one of the major health issues with an increased mortality rate. Cancer cells are distinguished by diverse genetic variations, epigenetic changes, and precarious chromosomal changes (Heng *et al.,* 2020). Bioinformatics and computational biology approaches can discern the key drivers of cancer. Nowadays, there is a focus on artificial intelligence-based approaches in personalized medicine in cancer therapy, paying the way for target-specific drugs to kill or disable diseased cells (Bhinder *et al.,* 2021). The latest cancer cases are estimated to rise to 19.3 million annually by 2025 globally. Social and economic shifts have contributed to increased chances of cancers associated with dietary and hormonal factors.

Environmental exposure, assaults, other supplementary factors, and chromosomal anomalies may often contribute to abnormalities in gene expression, including the level of expression perturbation in cell signalling and metabolic pathways (Heindel *et al.,* 2017). Traditional remedies have been useful in cancer treatment; they may have adverse side effects due to their nonspecific action on normal and tumor cells. Thus, targeted treatment strategies using small molecule inhibitors are widely studied. A druggable molecule's low molecular weight and small size may allow them to translocate via the plasma membrane to interact with intracellular receptors, cell-surface receptors, and intracellular signalling molecules. Small molecule inhibitors can be devised to target any part of the cell (Miersch *et al.,* 2016). Bioinformatics and computational modelling approaches ease the development of drugs based on-target specificity. Small molecule cancer drugs are smaller (≤500Da) and can translocate through plasma membranes. Various approaches are available to discover and confirm drugs for

specific targets for infectious diseases and bugs (Hassan *et al.,* 2021). Advances in bioinformatics and computational biology have amplified productivity at many phases of drug discovery (Atanasov *et al.,* 2021).

Natural products (such as plant metabolites, animal toxins, and immunological components) comprise an enormous and diverse source of bioactive compounds which have chemo-preventive potential, some of which are documented by age-old practices of traditional medicine and are largely separate from the set of small molecules usually used for drug discovery. The anticancer properties of bioactive compounds have been extensively studied, aiming for their application in some cancer treatment modalities, such as chemotherapy and targeted therapy (Majolo *et al*., 2019; Efferth *et al.,* 2021). Bioactive compounds can also participate as adjuvants in cancer chemotherapy and relieve undesirable effects (e.g., reactive oxygen species (ROS) scavenging) (Lin *et al.,* 2020). Huang *et al.*, demonstrated that resveratrol treatment significantly decreased downregulating the expression of (O-6-Methylguanine-DNA Methyltransferase) MGMT; it also decreased the expression of the antiapoptotic proteins, X-linked inhibitor of apoptosis protein (XIAP) and survivin. It has also been reported that resveratrol increased Temozolomide (TMZ) toxicity by increasing ROS generation, AMP-activated protein kinase (AMPK) pathway activation, and mechanistic target of rapamycin (mTOR) signalling inhibition and decreased antiapoptotic protein Bcl-2 expression in glioblastoma multiforme cells (Yuan *et al.*, 2012). Some bioinformatics approaches use supervised machine learning algorithms trained on protein structures (and motifs) with known activity to predict activity in new, uncharacterized proteins, which is essentially a traditional Quantitative structure-activity relationship (QSAR) designed to work on proteins (Trapotsi *et al.,* 2022). The FEATURE framework (Halperin *et al.,* 2008) uses the 3-dimensional spatial orientation of atoms to predict active sites in the macromolecular structure.

Artificial intelligence (AI) and machine learning applications in computer aided drug design had revolutionised the field of cancer medicine based on genomics, structural biology and combinatorial chemistry (Staszak *et al.,* 2022). Small molecule targeted therapy has entered in new phase with technological advancement in cancer biology. Thousands of molecules are going in clinical trials for various cancer treatments and FDA had approved 89 small molecules targeted cancer drugs (Zhong *et al.,* 2021). Small molecule based targeted therapy is facing biggest challenge of drug resistance over a period of time in clinical use. Drug resistance develops due to various mechanisms which include gene mutations, gene amplifications, expression of efflux transporters, autophagy, apoptosis dysregulation, epigenetic changes and stemness in some section of cells. The resistant mutant cells remodel the tumour microenvironment and creates havoc effects on normal cells and also are now under the regulation of drug resistant cells (Reagan *et al.,* 2019). Some mutations and amplifications like EGFR and MET amplification accounts a major section in resistance cases. Small molecule targeted therapy for cancer is facing another challenge of low efficacy due to the development of drug resistance and are effective only to limited number of patients (Seebacher *et al.,* 2019). Proteins that were previously thought as undruggable may also be a good anti-proliferative and anticancer target, one typical example is KRAS and mutated isoforms of RAS proto-oncogene (Moore *et al.,* 2020). It takes decades to unveil KRAS targets with small molecules because due to its strong binding affinity of its intrinsic ligand GTP and high flexibility switch regions (Chen *et al.,* 2020), Virtual screening of small molecules had developed that directly target KRAS via covalent allosteric binding, GTP-binding site switch pocket binding sites by protein-protein interactions (Nussinov *et al.,* 2022). Currently four KRAS G12C covalent inhibitors (AMG510, JNJ-74699157, MRTX849 and GDC-6036) are in clinical trials (Erlanson *et al.,* 2021). The epidermal growth factor receptor (EGFR) is a transmembrane protein implicated in a wide range of biological processes. Osimertinib is the first approved third-generation EGFR

inhibitor and can achieve a PFS of over 10 months in patients harbouring the EGFR T790M mutation (Remon *et al.,* 2018). Almonertinib, is an analog of Osimertinib (Afonso *et al.,* 2022). It also had showed significant anti-cancer effects in resistant patients with NSCLC in clinical trials, and has been approved for NSCLCtherapy by the NMPArecently (Liao *et al.,* 2022). Protein kinasesare the most widelystudiedtumortherapeutictargets. Anaplasticlymphomakinase (ALK) encoded by the *ALK* gene is a single transmembrane tyrosine kinase of the insulin receptor family. Currently, there are still some ALK inhibitors under clinical investigation, such as the pan-TKIs entrectinib, belizatinib, and repotrectinib, which target oncogenic rearrangements in ALK, ROS, and tropomyosin receptor kinase (TRK) (Zhong *et al.,* 2021). Among them, entrectinib was approved for the treatment of TRK fusion solid tumors in 2019. For resistance arising from bypass activation, combining ALK inhibitors with other targeted therapies such as a mitogen-activated protein kinase (MAPK) inhibitor, cyclin-dependent kinase (CDK) inhibitor (ceritinib with ribociclib, NCT02292550), mammalian target of rapamycin (mTOR) inhibitor (ceritinib with everolimus, NCT02321501), and heat-shock protein 90 inhibitor has been assessed in a number of trials. Since the expression of programmed death-ligand-1 (PDL-1) is reportedly associated with *EML4- ALK*, combined treatments of ALK and immune checkpoint inhibitors have also been evaluated in *ALK*-positive NSCLC (Hirsch *et al.,* 2016). Identifying new therapeutic uses for existing molecules which have benefits for patients who may not have efficacious therapy options. AI and machine learning provides a powerful avenue toward new research tools for interrogating disease biology and is an important window for drug repurposing. Drug repurposing identifying new therapeutic applications of existing or investigational drugs. Moreover, in some curious cases, inhibition of unintended targets may elicit unexpected beneficial effects in a different disease (Santos *et al.,* 2017), thus providing a foundation for drug repurposing. However, inhibition of a single target is often insufficient to generate optimal therapeutic benefit where the disease displays polygenicity (e.g., cancer,

psychiatric disorders) or involves complex biological signalling networks and feedback loops (A Antolin *et al.,* 2016; Workman *et al.,* 2013). For instance, receptor tyrosine kinases (RTKs) usually share the same canonical downstream signalling pathways, and inhibition of one RTK driver can be compensated for by activating a different RTK. Accordingly, EGFR-mediated bypass signalling confers resistance to anaplastic lymphoma kinase (ALK) or ROS1 inhibition by crizotinib in fusion-kinase-driven non-small cell lung cancer (NSCLC) (Davies *et al.,* 2013; Vaishnavi *et al.,* 2017), and MET signalling can counteract epidermal growth factor receptor (EGFR) inhibition by gefitinib in EGFR mutant NSCLC (Hata *et al.,* 2016). Similarly,feedback activation of EGFR in BRAF mutant colorectal cancer causes primary resistance to the BRAF inhibitor vemurafenib (Prahallad *et al.,* 2012). In the case of the PI3K–AKT mTORpathway, it is noteworthy that inhibition of mTORC1, for instance, by everolimus or knockdown of its downstream target S6K releases a negative feedback loop and paradoxically leads to upregulation of AKT signaling via, e.g., mTORC2 activation (Janku *et al.,* 2018; Rozengurt *et al.,* 2014).

Computer-aided drug design (CADD) has reduced the time and costs of developing new drugs based on clinically validated cancer drug targets using various drug repositories. Our goal is repositioning using molecular and quantum-based approaches to guide the development of new medications against multiple types of cancer. Moreover, CADD is an easy way to develop drugs based on drug simulation and molecular docking using the target-specific ligands. The docking validation is based on comparing the pose of a docked structure with its crystallographic pose, a process known as redocking. The crystallographic structures of ligands were extracted and reserved to compose the set of potentially active compounds at thevalidation step using the docking score and binding energy and based on the QSAR. Small molecules against various known validated targets will be studied for the development of novelmolecules by using molecular docking, and the interactions of small molecules and proteins

will be analyzed by STRING and Cytoscape. Our research work plan has frameworks based on AI and machine learning models, including deep learning models like convolutional neural networks.

2. **Aim:** Design of the novel therapeutic strategies in cancer(s) by employing *in silico* and *in vitro* studies.

3. Objectives

1. *In silico* screening of small molecule library against clinically validated target(s) of various cancers.
2. To identify the possible potent small molecule(s) against specific selected target(s) in various cancers.
3. *In vitro* validation of the selected molecule(s)-selective target(s)-specific cancer(s) for possible therapeutic strategies.

4. Review of Literature

Over 200 forms of cancer have been classified (Song *et al.,* 2015). Cancer cells have dynamic changes in the chromosomes, comprising various genetic aberrations and assaults such as chromosomal mutations, copy number variations, point mutations, modifications to gene expression profiles, and different epigenetic patterns. These alterations are diverse among different types of cancers (Jackson *et al.,* 2018). Traditional therapy which cytotoxically kill the rapidly dividing cells in the body by interfering with cell division without having specific target. Drug repositing have primary goal to find the drugs against the specific target using bioinformatic approaches to find novel inhibitors for specific clinically validated target(s). Novel molecules repositing is to halt the tumor growth and progression; Targeted therapy serves the best for tackling the cancer growth and progression with lowering potential side

effects. Targeted cancer molecules had been broadly classified either small molecule and macromolecules (monoclonal antibodies, polypeptides, antibody drug-conjugates and nucleic acids) (Zhong *et al.,* 2021). The basis of targets for various cancers are enzymes (kinases, proteases, esterase's and phosphatases), epigenetic inhibitors, proteosome inhibitors, structural proteins (tubulin), Poly (ADP-ribose) polymerases (PARPs), Nucleic acid targets (topoisomerases, intercalating drugs or alkylators), ion channels, hormones, growth factor receptors and other cell surface receptors. Enzymes serves good target for halting the tumor growth and progression (Jang *et al.,* 2013). Enzymes like kinases which regulate cell cycle, signaling pathways and chromatin remodelers, can reprogram the cancer cell by regulating cell proliferation, differentiation, migration, survival and death (Sever *et al.,* 2015). The dysregulation of various kinases (pyruvate kinase) may cause allergy, inflammatory disease, diabetes and auto immune diseases in pathological condition which triggers the tumor formation in chronic conditions (Patel *et al.,* 2021). Hormones and growth factors (GFs) had wide array of molecular functions that impact organismal development and can also have impact on pathophysiological state of an individual by regulating glucose and lipid metabolism (Nakagawa *et al.,* 2018). Hormones and GFs act as a signaling molecule implicated in the regulation of variety of cellular processes, they paly important role in both normal and tumor cells where they function by binding to specific receptors on a target cell to activate downstream signaling cascade. Dysregulation in the signaling pathway may advance the tumor growth and progression. Most of the tumor cells don't rely on the neighboring cells for GF synthesis the have own intrinsic mechanism for the synthesis of growth factors and they independently mediate their own metabolism and proliferation (Lappano *et al.,* 2022). The ectopic secretion hormones and GFs are associated with tumors of endocrine organs such as pancreatic islets cells, parathyroid cells thyroid cells and neuro-endocrine cells that can be found in every organ of humans. Hormones are classified into several groups among them the

steroid hormones (Estrogens and Androgens) were find in progression of glioblastoma (Bello *et al.,* 2021). they are considered risk factors for many malignancies such as breast, endometrial, ovarian, prostrate and thyroid tumors. Estrogen regulates various biological processes mainly by binding to the intracellular receptors such as ERα and Erβ which function as a ligand activated transcription factor that induce transcription of various target genes (Thomas *et al.,* 2011). Estrogen induces rapid response such as an increase in calcium and nitric oxide levels and also activation of certain kinases (Chen *et al.,* 2004). GFs have several groups such as epidermal growth factor (EGFR), neuregulin's (NGRs), Insulin-like growth factors, Tumor growth factor-β and vascular growth factors (VEGFs) which plays important role in cancer proliferation and progression (Witsch *et al.,* 2010).

Applying bioinformatics approaches is often critical in detecting and identifying the type of cancer characterized by a different molecular pattern that demands better therapeutic strategies. Human genetic makeup affects our likelihood of developing or resisting various diseases. Bioinformatics is executed within translational drug discovery in infectious diseases. For instance, viral or bacterial infection gives rise to specific gene expression profiles within the cell. Comparing these profiles with those of other conditions and drug-induced genetic profiles offers repositioning opportunities for existing drugs (Kim *et al.,* 2019). Since the human genome was first sequenced, genomic, proteomic, and metabolomic high-throughput platforms have increasingly allowed analyses of large datasets across many different diseases. Data science, machine learning, and statistical approaches are used to identify abnormal patterns that correlate with the disease process, often aiming to identify actionable targets that are druggable. Within any tumor, only a minority of the genetic changes enable and drive the progression of the disease. The other mutations provide no growth advantage and are often described as passenger mutations. Vogelstein *et al.*, identified approximately 140 potential genes that act as drivers of cancer. Many targeted anticancer drugs work by directly inhibiting activated

oncogenes, particularly proteins that contain protein kinase domains or proteins that are nuclear receptors (Yang *et al.,* 2021). For example, dabrafenib has been approved to treat late-stage melanoma and targets the constitutively activated kinase oncogene BRAF V600E. Cetuximab, panitumumab, gefitinib, and erlotinib are the licensed inhibitors of the EGFR tyrosine kinase, and crizotinib is an ALK inhibitor, all of which are approved for the treatment of lung cancer (Chitturi *et al.,* 2022).

Bioinformatics and computational biology approaches can discern the key drivers of cancer. Nowadays, there is a focus on virtual screening-based approaches in personalized medicine in cancer therapy, paying the way for target-specific drug designing to kill or disable the diseased cells (Bhinder *et al.,* 2021). The small-molecule inhibitor compounds are usually ≤500Da in size and are radially permeable to the plasma membrane. Small molecule inhibitors can be designed to target any part of the cell, regardless of the target's cellular location. The rise of multi-omics approaches unveiled various disease mechanisms that led to multitudes of ways to assess the outcome of the putative molecule(s) or compounds on cancer cells (Romano *et al.,* 2019). Although the size and complexity of proteins are often unaffordable to structure-based analyses designed for small molecules, other drug discovery approaches to leverage the distinctive characteristics of proteins and other macromolecules to perform discovery, bioinformatics, and computational modeling approaches ease ways for the development of drugs based on-target specificity. The leading strategies of specific molecular targeting known in clinical practice are small-molecule inhibitors and monoclonal antibodies (mAbs) targeting immune checkpoint inhibitor (Singh *et al.,* 2022). The small molecule inhibitors are also comparatively cost-effective and are amenable to oral administration, while the mAbs are mostly administered intravenously (Mardis *et al.,* 2008). Various small molecule inhibitors attach to numerous molecular targets, including cell surface receptors and other intracellular proteins, thus increasing the risk of toxicity (Moffat *et al.,* 2014). Most small molecule

inhibitors have a brief life span, thus requiring daily dosing (Zheng et al., 2013). Certain small molecule inhibitors have adverse side effects with following treatment have also been documented. For instance, treatment with imatinib can cause edema, diarrhea, vomiting, and nausea and sometimes lead to thrombocytopenia, anemia, and neutropenia (Eggert *et al.,* 2013). Drugs that impede angiogenesis can affect the formation of new blood vessels throughout the body, leading to detrimental consequences such as bleeding that becomes hard to operate (Muellner *et al.,* 2015). Drugs aginst EGFR can target the skin cells, which oftentimes express ample of EGFR, therefore inhibiting the normal growth of skin cells and driving it harder for the skin to maintain moisture, resulting in scaly and itchy skin (Castoreno *et al.,* 2010). Nilotinib, used in treating chronic myelogenous leukemia, carries a warning for possible heart complications (Gazdar *et al.,* 2010). The development of resistance to these drugs is another major drawback, and cancer cells acquiring resistance to drugs such as crizotinib and vemurafenib have been reported (Barretina *et al.,* 2012). BRAF inhibitors Vemurafenib and dabrafenib, although they have revolutionized the treatment of melanoma, approximately a fifth of the patients develop cutaneous squamous cell carcinoma (cSCC) during therapy. A recent report revealed that these BRAF inhibitors suppress apoptosis through their off-target inhibition of pro-apoptotic JNKs while concomitantly and paradoxically activating ERKs (Basu *et al.,* 2013). Various approaches are available to discover and confirm drugs for specific targets for infectious diseases and bugs (Agamah *et al.,* 2020). Epidermal growth factor receptor (EGFR) is also known as ErbB1, a transmembrane protein involved in numerous biological processes. Aberrant activation of EGFR driven by mutational changes which is a crucial contributor to the tumor formation and its heterogeneity leads to multiple forms of cancer, such as breast cancer, lung cancer, and pancreatic cancer (Roskoski *et al.,* 2014). The first generation of epidermal growth factor receptor tyrosine kinase inhibitors (EGFR-TKIs) is erlotinib, gefitinib, and icotinib, reversible in action with a quinazoline structure (Sun *et al.,*

2021). These inhibitors are highly effective in non-small cell lung cancer (NSCLC) carrying EGFR-activating mutations (exon 19 deletion and exon 21 L858R). The second-generation irreversible EGFR-TKIs are afatinib and dacomitinib (Lavacchi *et al.,* 2019). Pyrimidine-based third-generation EGFR TKIs contain inhibitory impacts on EGFR-activating mutations and the T790M (Ricordel *et al.,* 2018). Osimertinib is the first FDA-approved third-generation EGFR inhibitor devloped by Astrazeneca and can achieve progression-free survival of over 10 months in patients carrying the EGFR T790M mutation (Yver *et al.,* 2016). Almonertinib is an analog of osimertinib showed a significant effect on cancer-resistant with NSCLC in clinical trials, and it has also been approved for NSCLC treatment by the National Medicinal Product Administrationof China (Yang *et al.,* 2020).

Angiogenesis is a process occurring by a prominent pro-angiogenic factor known as vascular endothelial growth factor (VEGF), apart from VEGF, there are various factors also help in angiogenesis like basic fibroblast growth factor (bFGF), epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factor (TGF), insulin-like growth factor, and angiopoietin. Pemigatinib (INCB054828), is an orally potent FGFR selective inhibitor. pemigatinib got approval in April 2020 for the therapy of previously treated patients (Hoy *et al.,* 2020). It acts as a strong therapeutic molecule having advanced/metastatic or surgically unresectable cholangiocarcinoma harboring FGFR2 fusions or other rearrangements.179 Avapritinib is an effective and selective inhibitor of PDGFRα and KIT activation loop mutants, which has newly been approved in the United States for adults with unresectable or metastatic GIST bearing PDGFRα exon 18 mutations (Dhillon *et al.,* 2020).

Protein kinase is an enzyme that selectively catalyzes the transfer of γ- phosphate group from ATP to amino-acid residues containing hydroxyl groups in the protein molecule (Hirsch *et al.,* 2007). It plays a vital role in cell growth, proliferation, and differentiation. Aberrant activation

of protein kinases is linked to various diseases, specifically cancer. Anaplastic lymphoma kinase (ALK) is a single transmembrane tyrosine kinase constitutive activation of ALK through point mutations or chromosomal rearrangements has been identified in multiple human cancers such as anaplastic large cell lymphoma and diffuse large B-cell lymphoma (DLBCL) (Hallberg *et al.,* 2013), inflammatory myofibroblastic tumors, and non-small cell lung cancer (NSCLC). Crizotinib, approved in 2011, is a first-generation ALK inhibitor targeting multiple tyrosine kinases (Ou *et al.,* 2011), including ALK, cellular- mesenchymal-epithelial transition factor (c-Met), and proto-oncogene tyrosine-protein kinase reactive oxygen species (ROS). Presently, some ALK inhibitors are under clinical investigation, such as the pan-TKIs belizatinib, entrectinib, and repotrectinib (Drilon *et al.,* 2018), which target chromosomal translocation in ALK, and tropomyosin receptor kinase (TRK) (Menichincheri *et al.,* 2016), Entrectinib was first approved for TRK fusion, solid tumors in 2019 (Al-Salama *et al.,* 2019). The clinical usage of such drugs as ALK inhibitors is still under evaluation (Drilon *et al.,* 2017). The aminopyrazine derivative ensartinib is a recently formulated second-generation ALK inhibitor (Imran *et al.,* 2021). A phase III study is moving to compare ensartinib with crizotinib for the first-line treatment of *ALK*-positive NSCLC (NCT02767804) (Rothenstein *et al.,* 2018). CEP-37440 is an orally given inhibitor of ALK and focal adhesion kinase. This inhibitor has completed a phase I trial (NCT01922752) in patients with solid tumors (Infante *et al.,* 2012). In addition to kinase inhibitors, impairing carcinogenic proteins using proteolysis targeting chimera (PROTAC) technology is a productive anti-cancer strategy. The PROTACs MS4077 and MS4078 developed by Zhang et al. have shown great potency in lowering ALK fusion protein in preclinical studies, offering a new direction to drug discovery targeting ALK (Zhang *et al.,* 2018).

Fig1: Time line for the approval of small molecule targeted anti-cancer drugs (Source: signal transduction and targeted therapy, 2021 May 31;6(1):1-48).

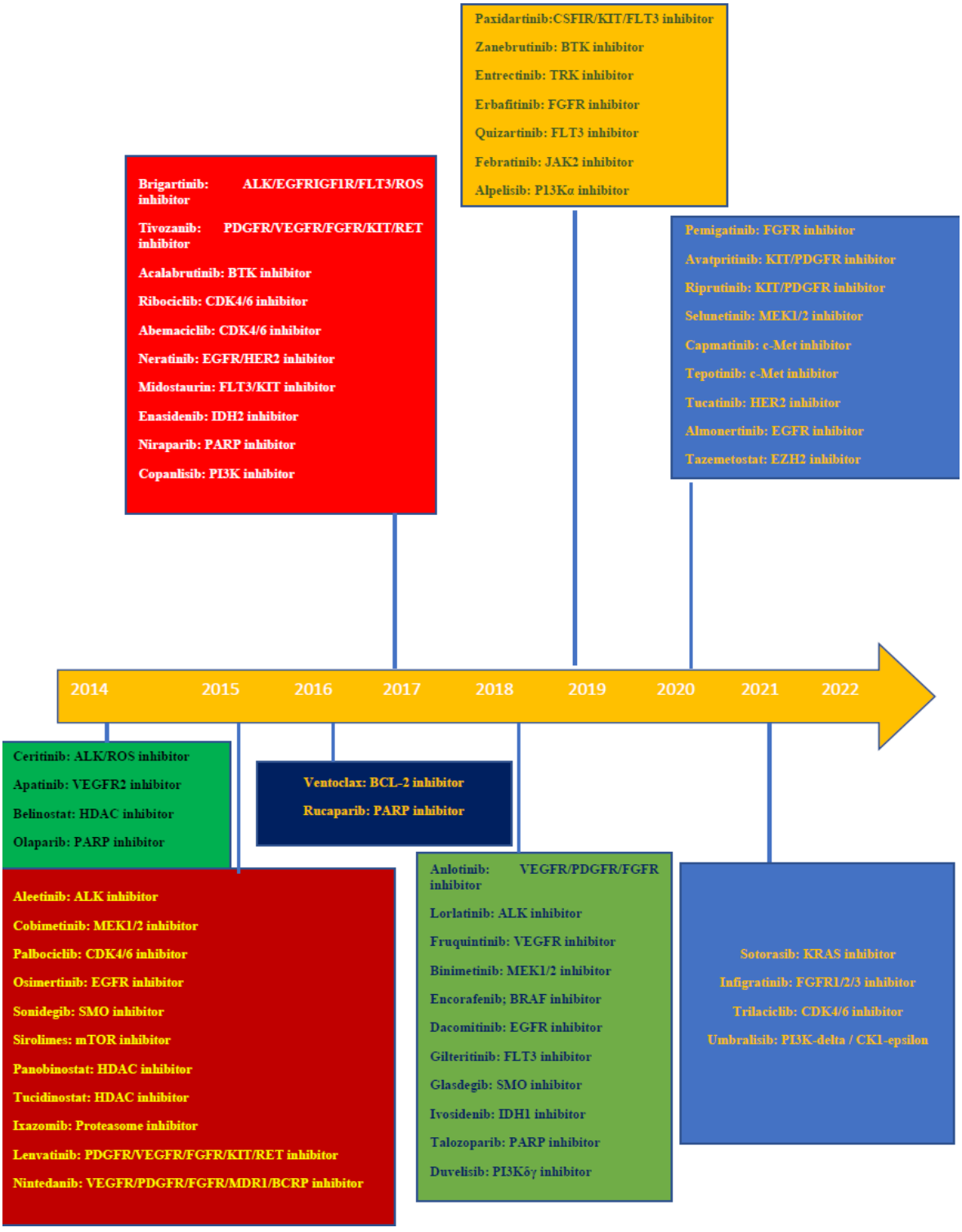

Fig2: Time line for the approval of small molecule targeted anti-cancer drugs (Source: signal transduction and targeted therapy, 2021 May 31;6(1):1-48).

Proteosome is a multi-enzyme subunit complex that degrades 80% of proteins in human cells, located in the cytoplasm and the nucleus. The ubiquitin-proteosome system (UPS) plays a vital role in maintaining homeostasis and removing misfolded proteins. The UPS is related to several human illnesses, such as cancers, autoimmune disorders, and genetic disorders; Multiple myeloma (MM) cells produce many paraproteins, and their growth depends on proteasome-regulated signaling pathways. Bortezomib is the first approved proteasome inhibitor (PI) for treating relapsed or refractory MM (Adams *et al.,* 2004). IT has also received FDA approval for the treatment of MCL. More than 200 clinical trials coupled to bortezomib concentrate on its combination with other agents, efficacy in other cancers, and even noncancer applications such as graft-versus-host disease. Marizomib can irreversibly bind to 3 major catalytic sites on β5, β1, and β2 of proteasome and interfere with proteasome activity at nanomolar concentrations (Groll *et al.,* 2006). The efficacy of marizomib in relapsed or refractory MM and lymphoma is evaluated in phase I/II trials (NCT00461045, NCT00396864) (Harrison *et al.,* 2016).

Dysfunction of cell cycle abnormalities results in uncontrolled cell proliferation. They have been regarded as one of the essential hallmarks of cancer (Hanahan *et al.,* 2011) Cyclin-dependent kinases (CDKs) are vital enzymes controlling/regulating the cell cycle, which requires cyclin proteins for activation and downstream phosphorylation. CDK4 and CDK6 are required to regulate growth signaling, which drives the cell cycle's transition from G1 to S. CDK4, and CDK6 are activated by cyclin-D, which induces the phosphorylation of tumor suppressor retinoblastoma protein 1 (RB1) in early G1 phase. CDK4/6-RB1 axis in mediating cellular proliferation and tumorigenesis, inhibition of CDK4/6 is an encouraging approach for cancer therapy that can induce cell cycle arrest in the G1 phase and result in reduced cell viability (Asghar *et al.,* 2015). Dysregulation of the CDK4/6-RB1 pathway is a significant characteristic of hormone receptor (HR)-positive breast cancer (Chen *et al.,* 2019). However,

three CDK inhibitors, palbociclib (Laderian *et al.,* 2017), ribociclib (Infante *et al.,* 2016), and abemaciclib to date are clinically available (Patnaik *et al.,* 2016). They are approved for the treatment of metastatic HR-positive, HER2-negative breast cancer in combination with specific endocrine therapies (Spring *et al.,* 2017). Despite existing approved only for specific breast cancer, the these approved CDK4/6 inhibitors have also been explored in other solid tumors and hematologic malignancies, such as lung cancer, prostate cancer, melanoma, glioblastoma, and myelofibrosis, and have shown promising results in preclinical studies or clinical trials of some malignancies (Tripathy *et al.,* 2017). Moreover, various newly formulated CDK4/6 inhibitors have joined clinical trials. Trilaciclib (G1T28) inhibits CDK4 and CDK6 with IC50 values of 1 and 4 nM, respectively. It has been assessed clinically to prevent chemotherapy-induced myelo suppression in triple-negative breast cancer (TNBC) and SCLC (He *et al.,* 2017). The pyridopyrimidine derivative PF-06873600 is potent against CDK2/4/6. It can overcome palbociclib resistance in preclinical studies. Phase II trials of PF-06873600 in treating metastatic breast cancer and other gynecological cancers are ongoing (NCT03519178). These CDK4/6 inhibitors are currently being studied in phase I/II trials (NCT03791112, NCT04488107, and NCT04539496) (Sanchez *et al.,* 2019).

Enhancer of zeste homolog 2 (EZH2) can change the expression of proteins. EZH2 is a histone methyltransferase and is a functional component of the polycomb repressor complex 2 (PRC2), which also includes other partners, such as embryonic ectoderm development (EED), suppressor of zeste 12 (SUZ12), and histone-binding proteins RbAp46/48.413 PRC2 is one of the two core complexes of polycomb group proteins (PcGs) which plays an essential role in methylation of H3 by transferring methyl groups from S-adenosyl-L-methionine (SAM) to lysine 27 resulting in chromatin condensation and transcriptional silencing of specific genes. Aberrant activation of EZH2 is closely associated with tumorigenesis and progression. It also leads to various malignant tumors, including breast cancer, prostate cancer, endometrial

carcinoma, ovarian cancer, melanoma, as well as hematological malignancies, such as NHL, B-cell lymphoma, and T-cell ALL (Sashida *et al.,* 2017). Since 2012, multiple SAM-competitive EZH2 inhibitors have been documented, including EI1, EPZ005687, GSK126, GSK343, UNC1999, tazemetostat (EPZ6438), SHR2554, CPI-1205, DS-3201, PF-06821497, and HH2853 (NCT04390737) (Duan *et al.,* 2020). These compounds show more suitable selectivity for EZH2 or EZH1 compared with other methyl transferases. Most of them bind to both wild-type and mutant EZH2, particularly Y641 and Y646 mutations (Bodor *et al.,* 2013). Nowadays, various inhibitors (e.g., tazemetostat, SHR2554, CPI-1205, DS-3201, PF-06821497, and HH2853) have entered clinical trials to assess their clinical efficacy and safety in a variety of solid tumors or hematological malignancies. Among them, tazemetostat, developed by Epizyme and Eisai, is an oral competitive inhibitor of the SAM pocket of the EZH2 SET domain and received accelerated approval in January 2020 in the USA for the treatment of adults and adolescents aged ≥16 years with locally advanced or metastatic epithelioid sarcoma not suitable for complete resection. Tazemetostat is the first approved EZH2 inhibitor and the first therapy specifically for treating epithelioid sarcoma (Hoy *et al.,* 2020).

Histone deacetylases (HDACs) are important epigenetic regulators that remove the acetyl groups from the N-acetylated lysine residues of histones and various non-histone substrates. To date, 18 HDACs have been identified in mammals. They are characterized into 4 subfamilies: class I HDACs (HDACs 1, 2, 3, 8), class II HDACs (class IIa: HDACs 4, 5, 7, 9; class IIb: HDACs 6, 10), class III HDACs (Sirt1-7) and class IV HDAC (HDAC11) (Park *et al.,* 2020). Class I, II, and IV HDACs are Zn2+-dependent histone deacetylases and can be controlled by inhibitors (such as TSA and SAHA) that occupy the catalytic core of the zinc-binding site. Class III HDACs require nicotinamide adenine dinucleotide (NAD+) for their activity Park *et al.,* 2020; Li *et al.,* 2016). Unusual upregulation of HDACs has been

documented in various cancers. Such transformations alter the transcription of oncogenes and tumor suppressor genes, which are closely associated with cell proliferation, apoptosis, differentiation, migration, and cancer angiogenesis (Li *et al.,* 2016; Eckschlager *et al.,* 2017). Inhibition of HDACs has been confirmed to be a productive strategy for cancer therapy. To date, various HDAC inhibitors have been approved by the FDA are undergoing clinical trials tazemetostat is among one of them.

Advances in bioinformatics and computational biology have amplified productivity at many phases of drug discovery (Atanasov *et al.,* 2021). Natural products (such as plant metabolites, animal toxins, and immunological components) comprise an enormous and diverse source of bioactive compounds which have chemo-preventive potential, some of which are documented by age-old practices of traditional medicine and are largely separate from the set of small molecules usually used for drug discovery. Moreover, natural products possess a special feature that distinguishes them from chemically synthesized small molecule drug candidates; they require new approaches for evaluating their therapeutic potential; Since many classes of natural products (NPs) comprise proteins, these approaches can often be adapted to NP drug discovery. Some bioinformatics approaches use supervised machine learning algorithms trained onprotein structures (and motifs) with known activity to predict activity in new, uncharacterizedproteins, which is essentially a traditional Quantitative structure-activity relationship (QSAR)designed to work on proteins (Trapotsi *et al.,* 2022). The FEATURE framework (Halperin *et al.,* 2008) uses the 3-dimensional spatial orientation of atoms to predict activity at numerous "active sites" within a larger macromolecule. It is therefore generalizable to diverse proteins with conserved functional activity. Furthermore, protein functional modeling approaches rely on input variables that behave like "abstractions" of natural molecular characteristics, includingamino acid or DNA structure (along with sequence alignment algorithms) (Vyas *et al.,* 2012),ontology annotations (Mutowo *et al.,* 2016), and biomarker response (Frank et al., 2003).

Computer-aided drug design (CADD) has reduced the time and costs of developing new drugs based on clinically validated cancer drug targets using various drug repositories in the last decades by using chemical-molecular and quantum-based approaches to guide the development of new drugs against various types of cancer. Moreover, CADD is an easy way to develop drugs based on drug simulation and molecular docking using the target-specific ligands. The docking validation is based on comparing the pose of a docked structure with its crystallographic pose, a process known as redocking. The crystallographic structures of ligands were extracted and reserved to compose the set of potentially active compounds at the validation step using the docking score and binding energy and based on the QSAR. Small molecules against various known validated targets will be studied for the development of novel molecules by using molecular docking, and the interactions of small molecules and proteins will be analyzed by STRING and Cytoscape. In combination with public biological databases (GO and KEGG), developing a high-throughput detection technology would systematically explore a list of DEG throughout the genome and search through the related binding partners. Drug repurposing refers to identifying new therapeutic applications of existing or investigational drugs. It may even allow 'drug rescuing' of compounds that have failed in initial clinical trials. Identifying new therapeutic uses for existing molecules has obvious benefits for patients that may otherwise not have any efficacious therapy options, but it also provides a powerful avenue towards new research tools for interrogating underlying disease biology. Various computational or *in silico* approaches have been widely used to explore drug repurposing opportunities, from drug target identification and prediction of new indications to mechanistic investigation in a given presentation (Achenbach *et al.,* 2011). *In silico* approaches are heavily driven by large data, such as various omics data sets (e.g., gene expression), the chemical structure space, and protein interaction databases. Harnessing gene expression signature-based Connectivity Map approaches to identify cellular targets and activities of

existing compounds in new diseases and infections (Subramanian *et al.,* 2017). Other systems biology approaches towards interrogating complex datasets and identifying new drug indications use, for instance, 'network pharmacology', that harnesses drug-gene or drug-protein correlations (Fernández *et al.,* 2019). The DEMAND hybrid method combines gene expression analysis with network perturbation upon drug treatment to elucidate drug-target interactions (Woo *et al.,* 2015). The similarity ensemble approach (SEA), on the other hand, focuses on the chemical structures of small molecule drugs or ligands based on the hypothesis that ligands with similar structures are likely to have similar targets (Keiser *et al.,* 2009). This concept has been shown to successfully identify multiple new drug-target interactions, leading to rapid repurposing if a new target is associated with a specific disease.

In contrast, the Repurposing Drugs in Oncology (ReDO) project is pursuing a literature-based approach to find licensed non-cancer drugs, which may also have utility in cancer based on prior reports of anticancer activity or indirect evidence based on activity against cancer-related pathways (Pantziarka *et al.,* 2014). In a clinical translation of off-target-based repurposing in cancer, there is a large number of studies that report new drug repurposing. Most of these involve kinase inhibitors, kinases have been identified to drive various cancers and thus are considered highly relevant cancer targets, and kinase inhibitors are notoriously unselective (although there are exceptions) and therefore make particularly good candidates for off-target-based drug repurposing.

5. Identification of Research Gaps

Genetic and epigenetic alterations in cell signalling pathways that govern cell-cycle progression, cell growth, and apoptosis are common hallmarks of cancer, but the degree of change in these mechanisms varies between individual tumors and tumor types. Over the last couple of years, the familiarity of cancer cell signaling inhibitors has greatly improved, leading

to the development of various targeted therapies that have elicited profound effects on cancer cells via acting as an epidrug or an inhibitor of target-specific in signalling pathways, but, unfortunately, short-lived responses this is somewhat due to the heterogeneity in the cancer cells. *In silico* approaches are largely driven by data generated by various omics data sets (e.g., gene expression), the chemical structure space, and protein interaction databases. Harnessing gene expression signature-based Connectivity Map and other systems biology approaches towards interrogating complex datasets and identifying novel small molecules against the clinically validated cancer target from the existing drug libraries. The repurposing of small mole molecule inhibitors may be a window to reduce the economic burden on cancer drug designing; furthermore, *in silico* approaches exploit the drug database to find a novel drug against the clinically validated targets.

6. Methodology

6.1. Methodology for Objective 1: *In silico* screening of small molecule library against clinically validated target(s) of various cancers.

6.1.1. Selection of Protein Target and known inhibitors

The target protein will be retrieved from Protein Data Bank – PDB (www.rcsb.org). It is the database for the three-dimensional structure of proteins and nucleic acids. The data provided by this database is obtained by X-ray crystallography, NMR spectroscopy and Cryo-Electron Microscopy. The data will be viewed by Pymol. The structure of protein will be downloaded in PDB format. Established inhibitors for clinically validated targets will be explored using NCBI's PubChem compound database, Zinc database, and Drug Library. Small molecules with a molecular weight below 300 Da will be screened. The selection of the ligand will be virtually screened using the Binding database. The virtually screened molecules will be selected based on various parameters of the binding database like IC50 value, EC50, Kd.

6.2. Methodology for Objective 2: To identify the possible potent small molecule(s) against specific selected target(s) in various cancers.

6.2.1. Preparation of Protein

Targeted three-dimensional structural coordinates will be pre-processed using the Protein Preparation Wizard module in Biovia Discovery Studio by implying the parameters like assigning bond orders, zero-order bonds to metal atoms, selenomethionine to methionine conversion, filling absent hydrogen, capping termini, side chains, loops, and removing waters beyond 5 Å distance surrounding the co-crystallized ligand. Further, tautomerization and protonation states will be predicted in favor of ligand at pH 7.00. Lastly, the protein hydrogen bonds were optimized to renovate the overlying hydrogens.

6.2.2. Lead Compounds and Database preparation

All the compounds retrieved from NCBI's PubChem Database and a large chemical library of drug-like compounds – Zinc databases will be prepared using the Ligand preparation module of molecular modeling package with suitable parameters like optimization, ring conformation, 2D to 3D conversion, determination of protomers, tautomers, and ionization states at pH 7.0, along with partial atomic charges using OPLS_2005 force field.

6..2.3. Active Site Prediction and Receptor Grid Generation

Using Biovia Discovery Studio, the binding site of the receptor will be predicted. Based on the top-ranked site score, five potential active sites will be analyzed. Only that site will be taken for docking analysis with the highest site score and the presence of hydrophobic and charged amino acids. Using the Receptor Grid Generation module in Biovia Discovery Studio, the grid generation will be performed on prepared protein with the help of a site score from SiteMap.

The protein atoms will be fixed within the default parameters of the radii of Vander Waal's scaling factor of 1 Å with a partial charge cut-off of 0.25Å using the OPLS_2005 force field.

6.2.4. Molecular Docking

Firstly, all the selected compounds will be docked into the generated grid of prepared protein using Auto Dock Vina, Biovia Discovery Studio with default parameters. Glide score is used to rank the various poses of ligand in a complex with a receptor, where the higher negative values reveal strong binding interaction of protein-ligand

6.2.5 ADME and Toxicity profile

PK properties such as absorption, distribution, metabolism, excretion, and toxicity (ADMET) profiling of compounds will be determined using the pkCSM ADMET descriptors algorithm protocol and the Discover Studio 4.0 (DS4.0) software package. The prediction of genotoxicity used the OECD QSAR toolbox 4.1 software package (Organization for Economic Co-operation and Development, Paris, France) and Toxtree, Version 2.6.13 (Ideaconsult, Ltd., Sofia, Bulgaria). Each software is an open-source freely available *in silico* program that identify the chemical structural alerts (SA).

6.3. Methodology for Objective 3: *In vitro* validation of the selected molecule(s)-selected target(s) against specific cancer(s).

6.3.1. Cell culture and treatment

Cells lines will be taken from the ATCC and grown on culture media. Culture media comprises an appropriate energy source and compounds that regulate the cell cycle. A typical culture media includes amino acids, vitamins, inorganic salts, glucose, and serum as a source of growth and attachment factors, and the selected compounds will be treated on the cell lines.

6.3.2. Cell viability assay

The MTT assay determines cellular metabolic activity as cell viability, cytotoxicity, and proliferation standard. This colorimetric assay is based on diluting a tetrazolium dye (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide or MTT) to its insoluble purple formazan crystals by metabolically active enzymatic conversions. The viable cells contain NAD(P)H dependent oxidoreductase enzymes, which reduce the MTT to formazan. Cells will be seeded into a polystyrene 96 well plate at a density of 10000 cells per well, and some wells will be left untreated, which will be taken as controls; other wells will be exposed to various bacterial strains (*H. pylori*) and incubated for 24 hours to test the effect on cell viability using MTT assay. MTT Plate will be incubated in a humidified 5% CO2 atmosphere at 37ºC at different intervals. In the MTT assay, the untreated wells will be treated as controls, while treated wells will be taken as a case. After adding MTT solution prepared in PBS, pH 7.4 was added after two and half hours. After this time, MTT and media will be discarded without removing the purple formazan crystals and will be followed by the addition of MTT solvent to solubilize the formazan crystals; the emanating-colored solution will be quantified by measuring absorbance at 500-600 nanometres using a multi-well spectrophotometer; the untreated wells will be taken as a control in calculating percent cell viability. The darker the solution, the greater the number of viable or metabolically active cells. Before infection, the medium of adherent cells was replaced by a medium containing 0.5% FCS. 1.5×10^4suspension cells per well will be seeded in 96-well plates for 2 h.

6.3.3. Expression Analysis

6.3.3.1 Cell lysate preparation and Western blot analysis

Cells will be seeded at a particular density in 24 wells plate. The cells will be given a treatment of selected small molecule(s), followed by washing with PBS and centrifugation. The cell pellets will be resuspended in ice-cold lysis buffer with freshly added protease and

phosphatase inhibitor cocktail for 20-30 minutes at 4°C. The cells will be centrifuged at around 10000 pm, and then supernatant (cell lysate) will be collected and transferred into a new eppendorfs for further procedure. The Bradford assay will be employed to estimate protein concentration using a known standard concentration of bovine serum albumin (BSA). This will be followed by separating proteins on SDS polyacrylamide gel after loading equal amounts of protein samples. The samples will then be transferred to an activated PVDF membrane for immunoblotting. Following membrane blocking, the overnight incubation with the primary antibody at 4ºC will be done. The membrane following washing with Tween-20/Tris-buffered saline (TBST) will be incubated with a peroxidase-conjugated secondary antibody at room temperature. The membrane after washing with TBST will be visualized with specific substrate-enhanced chemiluminescence (ECL) spotting reagent. The X-ray films will be examined, and the optical densities of the bands will be analyzed by densitometry.

6.3.4. Colony-formation assay

Cells will be seeded into 6-well plates (300 cells per well), treated with a selected compound and incubated for 10 days until the colonies are large enough to be discerned. The cells will be fixed with methanol and stained with crystal violet and the number of colonies with more than 50 cells will be counted manually. Experiments will be performed in triplicate and repeated three times.

6.3.5. Apoptosis Assay and Cell Cycle analysis

To evaluate the role of small molecule inhibitors on cells (treated and control) will be subjected to flow cytometry. It is a powerful technique used to detect and quantify the level of apoptosis in a population of cells. It allows the study of all aspects of apoptosis from induction via surface receptors to the late stage where DNA fragmentation occurs. It detects apoptosis using annexin V binding phosphatidylserine residue normally located within the plasma membrane. Phosphatidylserine residue is located outside during apoptosis, so only cells that have decided

to die will be detected by annexin V. Cell cycle analysis will be done by flow cytometry using DNA-based dyes to measure the DNA content.

6.3.6. Cell invasion and migration assay

Transwell culture inserts (8 μm pore size), 24-well plates, Matrigel (Millipore Corporation, Darmstadt, Germany) and pipette tips will be pre-cooled at 4 °C. Matrigel will be prepared 1:2 ratio in serum-free DMEM. An appropriate amount of Matrigel will be added to the upper chamber of the Transwell plates for the invasion assay, while the plates without Matrigel in the upper chamber will be used for the migration assay. Then, the chambers will be incubated at 37 °C in a 5% CO_2 incubator for 4–8 h. The treated cells will be resuspended into a single-cell suspension and seeded into the upper chamber at 2×10^5 cells/well density. Subsequently, 200 ml of serum-free DMEM will be added to the upper chamber, and 750 ml of complete DMEM containing 10% fetal bovine serum will be added to the lower chamber. Then cells will be incubated at 37 °C for 24 h in a 5% CO_2 incubator. The inserts will be collected, and the cells will be fixed in 4% polyformaldehyde at room temperature (RT) for 20 minutes and then stained with crystal violet for 30 minutes. The non-invading cells remaining in the Matrigel layer or on the upper surface of the Transwell insert membranes will be removed with a cotton swab. To observe the cells that had invaded through the semipermeable Transwell insert membrane and that had attached to the lower surface of the membranes, the inserts will be placed upside-down under an inverted phase-contrast microscope. Subsequently, three microscopic fields of view will be randomly selected. All of the above experiments will be repeated at least three times, and the average number of cells per high-power field will be calculated and then statistically analyzed.